

RECEIVED

MAR 28 2003

TECH CENTER 1600/2900

ENTERED

1600

RAW SEQUENCE LISTING

DATE: 03/20/2003

PATENT APPLICATION: US/09/703,951A

TIME: 16:10:53

Input Set : A:\SD9951 IA SeqList.txt

Output Set: N:\CRF4\03202003\I703951A.raw

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:
6 (i) APPLICANT: Gillespie, Alison
7 Claeps, Brian O.
8 Chavez-Noriega, Laura Elena
9 Siegel, Robert
10 Elliott, Kathryn J.
13 (ii) TITLE OF INVENTION: DNA ENCODING HUMAN (AND (SUBUNITS
14 OF NEURONAL NICOTINIC ACETYLCHOLINE
15 RECEPTOR, CELLS TRANSFORMED THEREWITH,
16 AND RECOMBINANT CELL LINE EXPRESSING
W--> 17 A HUMAN (AND (SUBUNITS OF NEURONAL
W--> 18 NICOTINIC ACETYLCHOLINE RECEPTOR
20 (iii) NUMBER OF SEQUENCES: 18
22 (iv) CORRESPONDENCE ADDRESS:
23 (A) ADDRESSEE: Merck & Co., Inc.
24 (B) STREET: 126 E. Lincoln Avenue
25 (C) CITY: Rahway
26 (D) STATE: NJ
27 (E) COUNTRY: USA
28 (F) ZIP: 07065-0907
30 (v) COMPUTER READABLE FORM:
31 (A) MEDIUM TYPE: Floppy disk
32 (B) COMPUTER: IBM PC compatible
33 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
34 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
36 (vi) CURRENT APPLICATION DATA:
C--> 37 (A) APPLICATION NUMBER: US/09/703,951A
C--> 38 (B) FILING DATE: 01-Nov-2000
39 (C) CLASSIFICATION:
61 (vii) PRIOR APPLICATION DATA:
42 (A) APPLICATION NUMBER: US 08/487,596
43 (B) FILING DATE: 07-JUN-1995
46 (A) APPLICATION NUMBER: WO US94/02447
47 (B) FILING DATE: 08-MAR-1994
50 (A) APPLICATION NUMBER: US 08/149,503
51 (B) FILING DATE: 08-NOV-1993
54 (A) APPLICATION NUMBER: US 08/028,031
55 (B) FILING DATE: 08-MAR-1993
58 (A) APPLICATION NUMBER: US 07/938,154
59 (B) FILING DATE: 30-NOV-1992
62 (A) APPLICATION NUMBER: US 07/504,455
63 (B) FILING DATE: 03-APR-1990

RAW SEQUENCE LISTING

DATE: 03/20/2003

PATENT APPLICATION: US/09/703,951A

TIME: 16:10:53

Input Set : A:\SD9951 IA SeqList.txt

Output Set: N:\CRF4\03202003\I703951A.raw

```

65 (viii) ATTORNEY/AGENT INFORMATION:
66 (A) NAME: Kohli, Vineet
67 (B) REGISTRATION NUMBER: 37,003
68 (C) REFERENCE/DOCKET NUMBER: SD9951IA
70 (ix) TELECOMMUNICATION INFORMATION:
71 (A) TELEPHONE: 732-594-3889
72 (B) TELEFAX: 732-594-4720
75 (2) INFORMATION FOR SEQ ID NO: 1:
77 (i) SEQUENCE CHARACTERISTICS:
78 (A) LENGTH: 2277 base pairs
79 (B) TYPE: nucleic acid
80 (C) STRANDEDNESS: both
81 (D) TOPOLOGY: both
83 (ii) MOLECULE TYPE: cDNA
86 (ix) FEATURE:
87 (A) NAME/KEY: CDS
88 (B) LOCATION: 166..1755
89 (D) OTHER INFORMATION: /product= "ALPHA-2 SUBUNIT"
92 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
94 CAATGACCTG TTTTCTTCTG TAACCACAGG TTCGGTGGTG AGAGGAASCY TCGCAGAATC 60
96 CAGCAGAATC CTCACAGAAT CCAGCAGCAG CTCTGCTGGG GACATGGTCC ATGGTGCAAC 120
98 CCACAGCAA GCCCTGACCT GACCTCCTGA TGCTCAGGAG AAGCCATGGG CCCCTCCTGT 180
100 CCTGTGTTCC TGTCCTTCAC AAAGCTCAGC CTGTGGTGGC TCCTTCTGAC CCCAGCAGGT 240
102 GGAGAGGAAG CTAAGCGCCC ACCTCCCAGG GTCCTGGAG ACCCACTCTC CTCTCCCAGT 300
104 CCCACGGCAT TGCCGAGGG AGGCTCGCAT ACCGAGACTG AGGACCGGCT CTTCAAACAC 360
106 CTCTTCCGGG GCTACAACCG CTGGGCGCGC CCGGTGCCCA ACACTTCAGA CGTGGTGATT 420
108 GTGEGCTTTG GACTGTCCAT CGCTCAGCTC ATCGATGTGG ATGAGAAGAA CCAAATGATG 480
110 ACCACCAACG TCTGGCTAAA ACAGGAGTGG AGCGACTACA AACTGCGCTG GAACCCCGCT 540
112 GATTTTGGCA ACATCACATC TCTCAGGGTC CTTTCTGAGA TGATCTGGAT CCCCACATT 600
114 GTTCTCTACA ACAATGCAGA TGGGGAGTTT GCAGTGACCC ACATGACCAA GGCCCACCTC 660
116 TTCTCCACGG GCACTGTGCA CTGGGTGCCC CCGGCCATCT ACAAGAGCTC CTGCAGCATC 720
118 GACGTCACCT TCTTCCCCTT CGACCAGCAG AACTGCAAGA TGAAGTTTGG CTCCTGGACT 780
120 TATGACAAGG CCAAGATCGA CCTGGAGCAG ATGGAGCAGA CTGTGGACCT GAAGGACTAC 840
122 TGGGAGAGCG GCGAGTGGGC CATCGTCAAT GCCACGGGCA CCTACAACAG CAAGAAGTAC 900
124 GACTGCTGCG CCGAGATCTA CCCCACGTC ACCTACGCTT TCGTCATCCG GCGGCTGCCG 960
126 CTCTTCTACA CCATCAACCT CATCATCCCC TGCCTGCTCA TCTCCTGCCT CACTGTGCTG 1020
128 GTCTTCTACC TGCCCTCCGA CTGCGGCGAG AAGATCACGC TGTGCATTTT GGTGCTGCTG 1080
130 TCACTCACCG TCTTCCTGCT GCTCATCACT GAGATCATCC CGTCCACCTC GCTGGTCATC 1140
132 CCGCTCATCG GCGAGTACCT GCTGTTCACC ATGATCTTCG TCACCCTGTC CATCGTCATC 1200
134 ACCGTCTTCG TGCTCAATGT GCACCACCGC TCCCCAGCA CCCACACCAT GCCCCACTGG 1260
136 GTGCGGGGAG CCCTTCTGGG CTGTGTGCCC CGGTGGCTTC TGATGAACCG GCCCCACCA 1320
138 CCCGTGGAGC TCTGCCACCC CTTACGCCTG AAGCTCAGCC CCTCTTATCA CTGGCTGGAG 1380
140 AGCAACGTGG ATGCCGAGGA GAGGGAGGTG GTGGTGGAGG AGGAGGACAG ATGGGCATGT 1440
142 GCAGGTCATG TGGCCCCCTC TGTGGGCACC CTCTGCAGCC ACGGCCACCT GCACTCTGGG 1500
144 GCCTCAGGTC CCAAGGCTGA GGCTCTGCTG CAGGAGGGTG AGCTGCTGCT ATCACCCAC 1560
146 ATGCAGAAGG CACTGGAAGG TGTGCACTAC ATTGCCGACC ACCTGCGGTC TGAGGATGCT 1620
148 GACTCTTCGG TGAAGGAGGA CTGGAAGTAT GTTGCCATGG TCATCGACAG GATCTTCCTC 1680
150 TGGCTGTTTA TCATCGTCTG CTTCTGGGG ACCATCGGCC TCTTTCTGCC TCCGTTCTTA 1740
152 GCTGGAATGA TCTGACTGCA CCTCCCTCGA GCTGGCTCCC AGGGCAAAGG GGAGGGTTCT 1800

```

RAW SEQUENCE LISTING

DATE: 03/20/2003

PATENT APPLICATION: US/09/703,951A

TIME: 16:10:53

Input Set : A:\SD9951 IA SeqList.txt

Output Set: N:\CRF4\03202003\I703951A.raw

```

154 TGGATGTGGA AGGGCTTTGA ACAATGTTTA GATTTGGAGA TGAGCCCCAA GTGCCAGGGA 1860
156 GAACAGCCAG GTGAGGTGGG AGGTTGGAGA GCCAGGTGAG GTCTCTCTAA GTCAGGCTGG 1920
158 GGTGAAGTT TGGAGTCTGT CCGAGTTTGC AGGGTGCTGA GCTGTATGGT CCAGCAGGGG 1980
160 AGTAATAAGG GCTCTTCCGG AAGGGGAGGA AGCGGGAGGC AGGGCCTGCA CCTGATGTGG 2040
162 AGGTACAGGG CAGATCTTCC CTACCGGGGA GGGATGGATG GTTGGATACA GGTGGCTGGG 2100
164 CTATTCCATC CATCTGGAAG CACATTTGAG CCTCCAGGCT TCTCCTTGAC GTCATTCCTC 2160
166 TCCTTCCTTG CTCCAAAATG GCTCTGCACC AGCCGGCCCC CAGGAGGTCT GGCAGAGCTG 2220
168 AGAGCCATGG CCTGCAGGGG CTCCATATGT CCCTACGCGT GCAGCAGGCA AACAAGA 2277
171 (2) INFORMATION FOR SEQ ID NO: 2:
173 (i) SEQUENCE CHARACTERISTICS:
174 (A) LENGTH: 529 amino acids
175 (B) TYPE: amino acid
176 (D) TOPOLOGY: unknown
178 (ii) MOLECULE TYPE: protein
181 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
183 Met Gly Pro Ser Cys Pro Val Phe Leu Ser Phe Thr Lys Leu Ser Leu
184 1 5 10 15
186 Trp Trp Leu Leu Leu Thr Pro Ala Gly Gly Glu Glu Ala Lys Arg Pro
187 20 25 30
189 Pro Pro Arg Ala Pro Gly Asp Pro Leu Ser Ser Pro Ser Pro Thr Ala
190 35 40 45
192 Leu Pro Gln Gly Gly Ser His Thr Glu Thr Glu Asp Arg Leu Phe Lys
193 50 55 60
195 His Leu Phe Arg Gly Tyr Asn Arg Trp Ala Arg Pro Val Pro Asn Thr
196 65 70 75 80
198 Ser Asp Val Val Ile Val Arg Phe Gly Leu Ser Ile Ala Gln Leu Ile
199 85 90 95
201 Asp Val Asp Glu Lys Asn Gln Met Met Thr Thr Asn Val Trp Leu Lys
202 100 105 110
204 Gln Glu Trp Ser Asp Tyr Lys Leu Arg Trp Asn Pro Ala Asp Phe Gly
205 115 120 125
207 Asn Ile Thr Ser Leu Arg Val Pro Ser Glu Met Ile Trp Ile Pro Asp
208 130 135 140
210 Ile Val Leu Tyr Asn Asn Ala Asp Gly Glu Phe Ala Val Thr His Met
211 145 150 155 160
213 Thr Lys Ala His Leu Phe Ser Thr Gly Thr Val His Trp Val Pro Pro
214 165 170 175
216 Ala Ile Tyr Lys Ser Ser Cys Ser Ile Asp Val Thr Phe Phe Pro Phe
217 180 185 190
219 Asp Gln Gln Asn Cys Lys Met Lys Phe Gly Ser Trp Thr Tyr Asp Lys
220 195 200 205
222 Ala Lys Ile Asp Leu Glu Gln Met Glu Gln Thr Val Asp Leu Lys Asp
223 210 215 220
225 Tyr Trp Glu Ser Gly Glu Trp Ala Ile Val Asn Ala Thr Gly Thr Tyr
226 225 230 235 240
228 Asn Ser Lys Lys Tyr Asp Cys Cys Ala Glu Ile Tyr Pro Asp Val Thr
229 245 250 255
231 Tyr Ala Phe Val Ile Arg Arg Leu Pro Leu Phe Tyr Thr Ile Asn Leu
232 260 265 270

```

RAW SEQUENCE LISTING

DATE: 03/20/2003

PATENT APPLICATION: US/09/703,951A

TIME: 16:10:53

Input Set : A:\SD9951 IA SeqList.txt

Output Set: N:\CRF4\03202003\I703951A.raw

```

234   Ile Ile Pro Cys Leu Leu Ile Ser Cys Leu Thr Val Leu Val Phe Tyr
235           275                      280                      285
237   Leu Pro Ser Asp Cys Gly Glu Lys Ile Thr Leu Cys Ile Ser Val Leu
238           290                      295                      300
240   Leu Ser Leu Thr Val Phe Leu Leu Leu Ile Thr Glu Ile Ile Pro Ser
241   305                      310                      315                      320
243   Thr Ser Leu Val Ile Pro Leu Ile Gly Glu Tyr Leu Leu Phe Thr Met
244           325                      330                      335
246   Ile Phe Val Thr Leu Ser Ile Val Ile Thr Val Phe Val Leu Asn Val
247           340                      345                      350
249   His His Arg Ser Pro Ser Thr His Thr Met Pro His Trp Val Arg Gly
250           355                      360                      365
252   Ala Leu Leu Gly Cys Val Pro Arg Trp Leu Leu Met Asn Arg Pro Pro
253           370                      375                      380
255   Pro Pro Val Glu Leu Cys His Pro Leu Arg Leu Lys Leu Ser Pro Ser
256   385                      390                      395                      400
258   Tyr His Trp Leu Glu Ser Asn Val Asp Ala Glu Glu Arg Glu Val Val
259           405                      410                      415
261   Val Glu Glu Glu Asp Arg Trp Ala Cys Ala Gly His Val Ala Pro Ser
262           420                      425                      430
264   Val Gly Thr Leu Cys Ser His Gly His Leu His Ser Gly Ala Ser Gly
265           435                      440                      445
267   Pro Lys Ala Glu Ala Leu Leu Gln Glu Gly Glu Leu Leu Leu Ser Pro
268   450                      455                      460
270   His Met Gln Lys Ala Leu Glu Gly Val His Tyr Ile Ala Asp His Leu
271   465                      470                      475                      480
273   Arg Ser Glu Asp Ala Asp Ser Ser Val Lys Glu Asp Trp Lys Tyr Val
274           485                      490                      495
276   Ala Met Val Ile Asp Arg Ile Phe Leu Trp Leu Phe Ile Ile Val Cys
277           500                      505                      510
279   Phe Leu Gly Thr Ile Gly Leu Phe Leu Pro Pro Phe Leu Ala Gly Met
280           515                      520                      525
282   Ile
286 (2) INFORMATION FOR SEQ ID NO: 3:
288   (i) SEQUENCE CHARACTERISTICS:
289       (A) LENGTH: 1654 base pairs
290       (B) TYPE: nucleic acid
291       (C) STRANDEDNESS: both
292       (D) TOPOLOGY: both
294   (ii) MOLECULE TYPE: cDNA
297   (ix) FEATURE:
298       (A) NAME/KEY: CDS
299       (B) LOCATION: 39..1553
300       (D) OTHER INFORMATION: /product= "ALPHA-3 SUBUNIT"
303   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
305   CCGACCGTCC GGGTCCGCGG CCAGCCCGGC CACCAGCCAT GGGCTCTGGC CCGCTCTCGC      60
307   TGCCCCTGGC GCTGTCGCCG CCGCGGCTGC TGCTGCTGCT GCTGTCTCTG CTGCCAGTGG      120
309   CCAGGGCCTC AGAGGCTGAG CACCGTCTAT TTGAGCGGCT GTTTGAAGAT TACAATGAGA      180
311   TCATCCGGCC TGTAGCCAAC GTGTCTGACC CAGTCATCAT CCATTTGAG GTGTCCATGT      240

```

RAW SEQUENCE LISTING

DATE: 03/20/2003

PATENT APPLICATION: US/09/703,951A

TIME: 16:10:53

Input Set : A:\SD9951 IA SeqList.txt

Output Set: N:\CRF4\03202003\I703951A.raw

```

313 CTCAGCTGGT GAAGGTGGAT GAAGTAAACC AGATCATGGA GACCAACCTG TGGCTCAAGC      300
315 AAATCTGGAA TGACTACAAAG CTGAAGTGGA ACCCCTCTGA CTATGGTGGG GCAGAGTTCA      360
317 TCGGTGTCCC TGCACAGAAAG ATCTGGAAGC CAGACATTGT GCTGTATAAC AATGCTGTTG      420
319 GGGATTTTCCA GGTGGACGAC AAGACCAAAG CCTTACTCAA GTACACTGGG GAGGTGACTT      480
321 GGATACCTCC GGCCATCTTT AAGAGCTCCT GTAAAATCGA CGTGACCTAC TTCCCGTTTG      540
323 ATTACCAAAA CTGTACCATG AAGTTCGGTT CCTGGTCCTA CGATAAGGCG AAAATCGATC      600
325 TGGTCCTGAT CGGCTCTTCC ATGAACCTCA AGGACTATTG GGAGAGCGGC GAGTGGGCCA      660
327 TCATCAAAGC CCCAGGCTAC AAACACGACA TCAAGTACAG CTGCTGCGAG GAGATCTACC      720
329 CCGACATCAC ATACTCGCTG WWCATCCGGC GGCTGTCGTT GTTCTACACC ATCAWCCTCA      780
331 TCATCCGCTG GCTGATCATC TCCTTCATCA CTGTGGTCGT CTCCTACCTG CCCTCCGACT      840
333 GCGGCGAGAA GGTGACCCTG TGYATTTCTG TCCTCCTCTC CCTGACGGTG TTTCTCCTGG      900
335 TGATCACTGA GACCATCCCT TCCACCTCGC TGGTCATCCC CCTGATTGGA GAGTACCTCC      960
337 TGWWCACCAT GATTTGTGTA ACCTTGTTCA TCGACATCAC CGTCTGCGTG CTCAACGTGC      1020
339 ACTACAGAAC CCCGACGACA CACACAATGC CCTCATGGGT GAAGACTGTA TTCTTGAMCC      1080
341 TGCTCCCCAG GGTCAATGTC ATGACCAGGC CAACAAGCAA CGAGGGCAAC GCTCAGAAGC      1140
343 CGAGGCCCTT CTACGGTGCC GAGCTCTCAA ATCTGAATTG CTTGAGCCGC GCAGAGTCCA      1200
345 AAGGCTGCAA GGAGGGCTAC CCCTGCCAGG ACGGGATGTG TGTTACTGTC CACCACCGCA      1260
347 GGATAAAAAT CTCCAATTTT AGTGCTAACC TCACGAGAAG CTCTAGTTCT GAATCTGTTG      1320
349 ATGCTGTGCT GTCCCTCTCT GCTTTGTGAC CAGAAATCAA AGAAGCCATC CAAAGTGTC      1380
351 AGTATATTGC TGAAAAATATG AAAGCACAAA ATGAAGCCAA AGAGATTCAA GATGATTGGA      1440
353 AGTATGTTGC CATGGTGATT GATCGTATTT TTCTGTGGGT TTTCACCCTG GTGTGCATTC      1500
355 TAGGGACAGC AGGATTGTTT CTGCAACCCC TGATGGCCAG GGAAGATGCA TAAGCACTAA      1560
357 GCTGTGTGCC TGCCTGGGAG ACTTCCTTGT GTCAGGGCAG GAGGAGGCTG CTTCTAGTA      1620
359 AGAACGTAAT TTCTGTTATC AAGCTACCAG CTTT                                1654

```

362 (2) INFORMATION FOR SEQ ID NO: 4:

364 (i) SEQUENCE CHARACTERISTICS:

365 (A) LENGTH: 504 amino acids

366 (B) TYPE: amino acid

367 (D) TOPOLOGY: unknown

369 (ii) MOLECULE TYPE: protein

372 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

374 Met Gly Ser Gly Pro Leu Ser Leu Pro Leu Ala Leu Ser Pro Pro Arg
375 1 5 10 15
377 Leu Leu Leu Leu Leu Leu Ser Leu Leu Pro Val Ala Arg Ala Ser Glu
378 20 25 30
380 Ala Glu His Arg Leu Phe Glu Arg Leu Phe Glu Asp Tyr Asn Glu Ile
381 35 40 45
383 Ile Arg Pro Val Ala Asn Val Ser Asp Pro Val Ile Ile His Phe Glu
384 50 55 60
386 Val Ser Met Ser Gln Leu Val Lys Val Asp Glu Val Asn Gln Ile Met
387 65 70 75 80
389 Glu Thr Asn Leu Trp Leu Lys Gln Ile Trp Asn Asp Tyr Lys Leu Lys
390 85 90 95
392 Trp Asn Pro Ser Asp Tyr Gly Gly Ala Glu Phe Met Arg Val Pro Ala
393 100 105 110
395 Gln Lys Ile Trp Lys Pro Asp Ile Val Leu Tyr Asn Asn Ala Val Gly
396 115 120 125
398 Asp Phe Gln Val Asp Asp Lys Thr Lys Ala Leu Leu Lys Tyr Thr Gly
399 130 135 140

```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 03/20/2003
PATENT APPLICATION: US/09/703,951A TIME: 16:10:54

Input Set : A:\SD9951 IA SeqList.txt
Output Set: N:\CRF4\03202003\I703951A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:4; Xaa Pos.235,246,309,347,354

Seq#:5; N Pos. 26,50,2278

VERIFICATION SUMMARY

DATE: 03/20/2003

PATENT APPLICATION: US/09/703,951A

TIME: 16:10:54

Input Set : A:\SD9951 IA SeqList.txt

Output Set: N:\CRF4\03202003\I703951A.raw

L:37 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:38 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:17 M:259 W: Allowed number of lines exceeded, (ii) TITLE OF INVENTION:
L:18 M:259 W: Allowed number of lines exceeded, (ii) TITLE OF INVENTION:
L:416 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4 after pos.:224
M:341 Repeated in SeqNo=4
L:1110 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:9